Model Assisted Qualification of NDE Techniques

Alexander Ballisat[1, 2, a)], Paul Wilcox[1)], Robert Smith[1)] and David Hallam[2)]

[1]*University of Bristol, Department of Mechanical Engineering, Queen's Building, University Walk, Bristol, BS8 1TR, United Kingdom*

[2]*Dstl, Porton Down, Salisbury, Wiltshire, SP4 0JQ, United Kingdom*

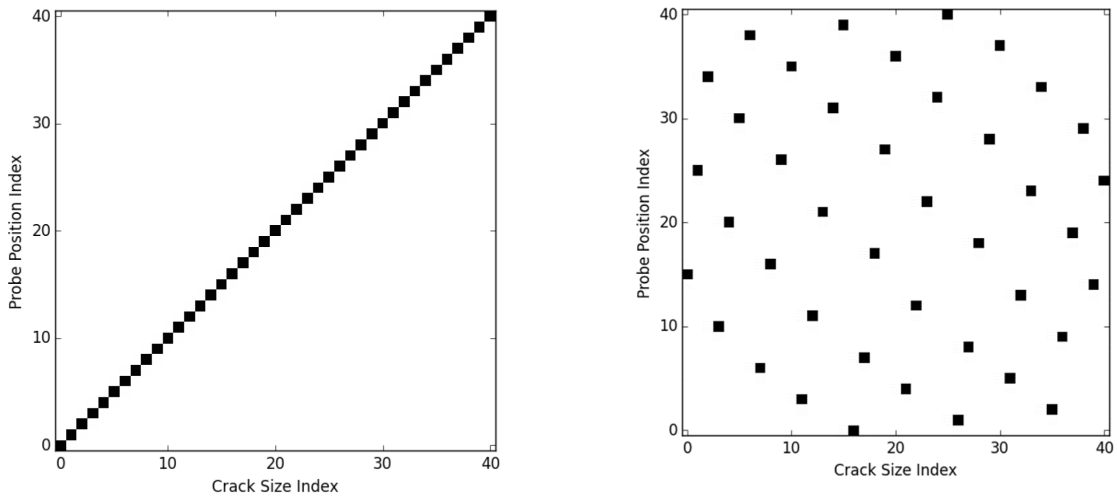
[a)]Corresponding Author: Alexander.Ballisat@bristol.ac.uk

Abstract. The costly and time consuming nature of empirical trials typically performed for NDE technique qualification is a major barrier to the introduction of NDE techniques into service. The use of computational models has been proposed as a method by which the process of qualification can be accelerated. However, given the number of possible parameters present in an inspection, the number of combinations of parameter values scales to a power law and running simulations at all of these points rapidly becomes infeasible. Given that many NDE inspections result in a single valued scalar quantity, such as a phase or amplitude, using suitable sampling and interpolation methods significantly reduces the number of simulations that have to be performed. This paper presents initial results of applying Latin Hypercube Designs and Multivariate Adaptive Regression Splines to the inspection of a fastener hole using an oblique ultrasonic shear wave inspection. It is demonstrated that an accurate mapping of the response of the inspection for the variations considered can be achieved by sampling only a small percentage of the parameter space of variations and that the required percentage decreases as the number of parameters and the number of possible sample points increases. It is then shown how the outcome of this process can be used to assess the reliability of the inspection through commonly used metrics such as probability of detection, thereby providing an alternative methodology to the current practice of performing empirical probability of detection trials.

INTRODUCTION

The qualification of NDE techniques for in-service use is a problem faced across a wide range of NDE sectors. Within the aerospace industry qualification of techniques is currently achieved through empirical trials involving a range of sample defects, equipment and operators. These trials are intended to demonstrate that given the variations present in an inspection, an operator is capable of consistently detecting defects reliably. These trials, as well as being time consuming and costly, rely on the samples being realistic representations of defects likely to be present in service which is often not a valid assumption and obtaining real samples from service is often difficult. The burden this process places on organizations has been identified as a key barrier to the introduction of techniques into service [1]. The use of experimentally validated computational models to replace a significant number of empirical trials has been proposed as a method through which this barrier may be significantly reduced. Primarily, this approach would significantly reduce the cost burden of qualification and the time required to bring new techniques into service. The use of computational models also has the benefit of being able to model realistic defects which will increase the accuracy of the qualification result. The use of models can also yield information beyond just the probability of detection, including the generation of receiver-operator characteristic curves which gives information on the likelihood of false calls, providing other metrics by which inspections can be assessed.

This paper presents initial results of using sampling and interpolation to greatly reduce the number of models required to obtain a measure of the reliability of an inspection by demonstrating these methods on the ultrasonic inspection of a fastener hole crack.



(a) (b)

Figure 1. Two Latin Hypercube Designs for example inspection variations with 41 points each. (a) The simplest Latin Hypercube Design which is the diagonal and (b) a design generated by the Translational Propagation method.

SAMPLING AND INTERPOLATION

In order to quantify the reliability of an inspection, its outcome given the parameters present must be known. Thus ideally simulations would be run at every combination of parameter values. This, however, scales to a power law and rapidly becomes infeasible as the number of parameters considered increases, especially if the model has a long run time. The outcome of many inspections, however, are single valued scalar quantities, such as amplitude or phase. Therefore, the response space which describes the outcome of the inspection for all of the parameters present can be sampled and these points used to interpolate over the entire space, greatly reducing the number of simulations that have to be performed. The challenges to achieve this are to choose an interpolation method which can fit to any features which may be present in the response space and to select a sampling method which when combined with the interpolation method can accurately map the response space using as small as possible number of simulations. Furthermore, a termination condition is needed to end the mapping process when sufficient convergence is achieved as it will not be possible to fully map the response space. Given these challenges, the response mapping algorithm proposed is as follows:

1. Generate an initial set of sample points in the parameter space of variations.
2. Perform simulations at these points.
3. Use the outputs of the initial sample set models to interpolate over the entire parameter space.
4. Generate a sample set independent of the initial sample set and perform simulations at those points.
5. Compare the outputs to the interpolated values at the same points generated in step 3.
6. If the error of this prediction is within a tolerance stop, otherwise use the combination of these two sample sets as the initial sample set in step 3 and repeat steps 3-6 to until the tolerance condition is met.

The design of experiments is a broad field with a significant body of literature. There are a wide range of experimental designs that could be appropriate as a sampling strategy and in this work Latin Hypercube Designs (LHDs) have been chosen as there is evidence [2] that these provide more robust estimates of functions for complex systems than random sampling. These designs are defined by each value of each variation being sampled once and only once. Two example designs for two variations with 41 values each are shown in Fig. 1. Figure 1a shows the simplest design which is sampling along the positive diagonal of the parameter space. The sampling in this design is entirely focused on the positive diagonal and interpolating the response over the parameter space using these samples would likely lead to an inaccurate result. Therefore a method is desired which generates a design with good sampling throughout the space in a short time. There are several metrics of sampling quality used in the literature and in this work the Φ_p [2] criterion is used, defined by the following equation:

$$\Phi_p = \left[\sum_{i=1}^{n-1} \sum_{j=1}^{n} d_{ij}^{-p} \right]^{-p} \tag{1}$$

where n is the number of points in the design, d is the Euclidean distance between points i and j. The best value of p is subject to debate and in this work $p = 50$ [3] is chosen as a compromise between maximizing the minimum distance between any two points ($p \rightarrow \infty$) and minimizing the total distance between all the points ($p = 1$). For the diagonal design in Fig. 1a $\Phi_p = 0.76$. Generating a highly optimal design quickly is not a trivial task. The number of possible designs for m parameters with n values each is $(n!)^{m-1}$ therefore brute force optimization of a design rapidly becomes impractical as the number of parameters and number of sample points increases, for example 41 points for two parameters has 3.35×10^{49} possible designs. The method used in this work to generate highly optimal designs is the Translational Propagation Algorithm [4] which uses the propagation of a small initial seed design to generate the final design. An example of the outcome of this algorithm is shown in Fig. 1b which has $\Phi_p = 0.21$. The comparison of the Φ_p values demonstrates that the LHD generated by the Translational Propagation method has better sampling properties than the diagonal design.

There is no prior knowledge as to what features may be present in the response space, such as discontinuities. It is also possible that there may be interaction between parameters. The Multivariate Adaptive Regression Splines interpolation method [5] is capable of handling such features and is briefly summarized here. The model that is fitted to the data is of the form given by the following equation:

$$g(\mathbf{X}) = \beta_0 + \sum_{i=1}^{M} \beta_i \prod_{j=1}^{n} \prod_{k=1}^{v} h(x_j, c_k) \tag{2}$$

where $\mathbf{X}$ is the input parameters, β_0 is the constant intercept term, β_i is the coefficient for the i^{th} basis function and each basis function is a product of hinge functions h. Each of the M basis functions are a product of nv hinge functions for n parameters with v knots each. The hinge functions are parameterized by a parameter x, knot location c and a direction, either positive or negative. The hinge function for a single parameter is defined by the following equation:

$$h(x,c)^{\pm} = \max(0, \pm(x-c))^{+} \tag{3}$$

The placement of knot locations allows for discontinuities in the response and the discontinuity in the first derivative at the knot location allows for discontinuities in the derivative of the response. The model is fitted through a recursive process in two stages. In the forward pass, reflected pairs of basis functions are added to the model in pairs and a least squares regression is performed of the model after each iteration. This process continues until a constraint on either the number of basis functions added or an error on the fit to the data is met. The outcome of this is typically an over constrained model thus a backward pass is subsequently performed. This removes added basis functions individually, fits the reduced model again and finds the optimal set of basis functions for the final model based on a cost function of the number of basis functions and the fit error of the model.

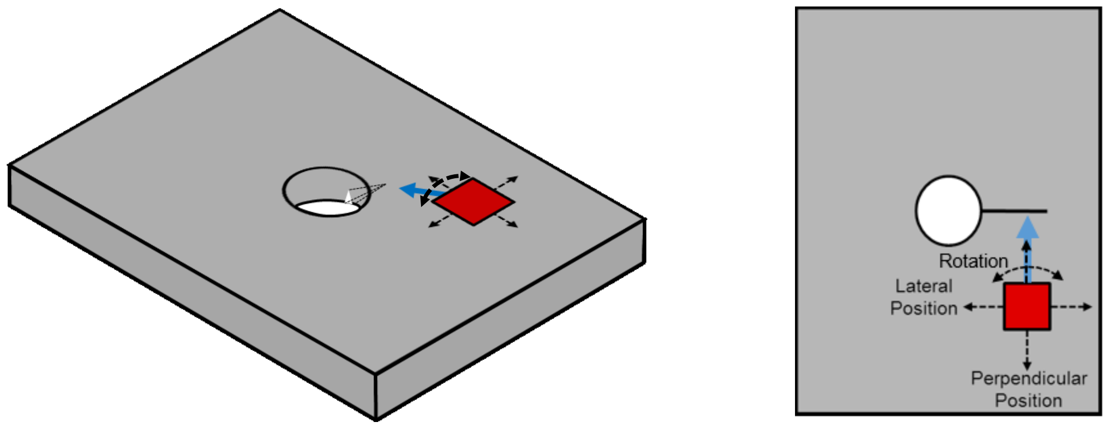
The model described in Equation 2 can be recast into the following form:

$$g(\mathbf{X}) = \beta_0 + \sum_{\sigma_i} \beta_i f(x_i) + \sum_{\sigma_{ij}} \beta_{ij} f(x_i, x_j) + \sum_{\sigma_{ijk}} \beta_{ijk} f(x_i, x_j, x_k) + \ldots \tag{4}$$

where the first sum is over basis functions of a single variable, the second is a sum over basis functions of two variables and so on. This can be used to perform an ANOVA decomposition on the function which yields information on the relative importance of parameters and the interaction between them. This can be used to judge whether a parameter has sufficiently little bearing on the outcome of the inspection that it could be removed from the parameter space, decreasing its dimensionality and reducing the size of the response space being mapped.

INSPECTION EXAMPLE

A canonical problem in aerospace NDE is the inspection of cracks emanating from fastener holes. This is commonly performed using a single element ultrasound probe utilizing an angled shear wave inspection. A simplified version of this problem was modelled using Pogo FEA [6]. The simulations were performed on a NVIDIA TitanX GPU (NVIDIA Corporation, California, USA) and each took approximately 20 minutes for pre-processing, solving and post-processing the output. The model is shown in Fig. 2 and consists of a 6.5 mm thick aluminum plate with a 6



(a) (b)

FIGURE 2. Schematic diagrams of the inspection investigated from a fastener hole: (a) isometric and (b) plan view.

mm diameter hole drilled in the center. A triangular shaped crack of height 2.89 mm, base width 0.1 mm and length 5 mm was placed on the side of the hole. A 45° shear wave 5 MHz ultrasound pulse was induced by applying an appropriately-phased pattern of normal surface traction to simulate the excitation of shear waves at 45° by a wedge probe on the surface. The inspection was assessed through a gated threshold method and sentenced by a response in the gate greater than the threshold being recorded as a positive indication and below the threshold being a negative indication. The following two sections present the results of allowing firstly two parameters to vary followed by the more complicated scenario of three parameters varying. Both of these scenarios have identical model inputs and the same mesh was used for all finite element calculations.

Two Parameter Scenario

In this scenario two parameters were allowed to vary: the lateral position and the perpendicular position relative to the longest crack axis. These are constrained to the ranges [4 mm, 9 mm] and [5 mm, 10 mm] for the lateral and perpendicular positions relative to the center of the hole respectively. The calculation of inspection metrics, such as probability of detection, requires a probability of variation to be associated with each parameter. The probability density function for the lateral position was modelled by a Gaussian of mean 6.5 mm and standard deviation 1 mm. Similarly, the probability density function for the perpendicular position was also modelled by a Gaussian with mean 7.5 mm and standard deviation 1 mm. It is assumed that the variations occur independently, thus to calculate the probability P of an inspection occurring at a given coordinate (x, y) in the parameter space, the following equation can be used:

$$P(x,y) = P(x)P(y) \quad 5)$$

Each parameter was assigned 11 possible values, giving a total parameter space of 121 points. This is a very small space which enables the space to be fully mapped to provide a benchmark solution against which interpolated results from sub-sampling can be compared. The simulated normalized maximum amplitude response map for this parameter space is shown in Fig. 3. It shows that there is a significant variation in the response amplitude as the positions change. Traditionally, probability of detection is calculated as a function of the defect size or some other defect parameter. In this example, the defect size is fixed, however the information generated in this method allows the probability of detection to be calculated as a function of any parameter considered. This is achieved by fixing the desired parameter at a range of values and using the variation in the response caused by the variations of the other parameters to calculate the probability of the response being above the threshold as the desired parameter is fixed at each value. If a traditional probability of detection curve was desired then one of the parameters could be the defect size, or any defect parameter which can be modelled, and the same process followed. Figure 4 shows the probability of detection curves for both the lateral and perpendicular positions. They indicate that the perpendicular position is a more significant factor than the lateral position as there is a larger variation in the probability of detection for that parameter. This can also be assessed by casting the fitted model in the form of Equation 4 and performing an ANOVA decomposition. The result

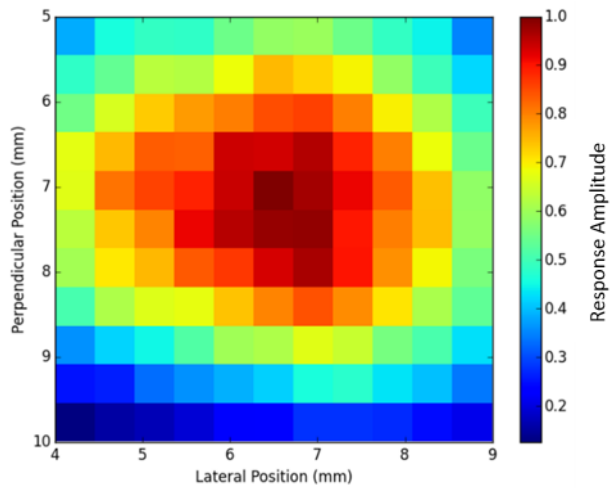


Figure 3. The normalized maximum amplitude response map for the parameter space in the two parameter scenario calculated using Pogo FEA.

of this is summarized in Table 1. This shows, as predicted from the probability of detection curves, that the perpendicular position is the most important parameter and that there is also interaction between the two parameters.

As it was possible to fully map the response space it is possible to compare the prediction to both the fully mapped space and the next set of samples generated in step 4 of the mapping algorithm. The result of this is shown in Fig. 5 which demonstrates that the error of the prediction compared to both the full response map and independent sample sets decreases to a low value rapidly. The sampling method approximately tracks the full map error. This suggests that the mapping algorithm is a valid method of predicting the response of the inspection. Whilst the percentage of points needed to map the space to a high degree of accuracy is relatively large, it is an improvement over mapping the whole space.

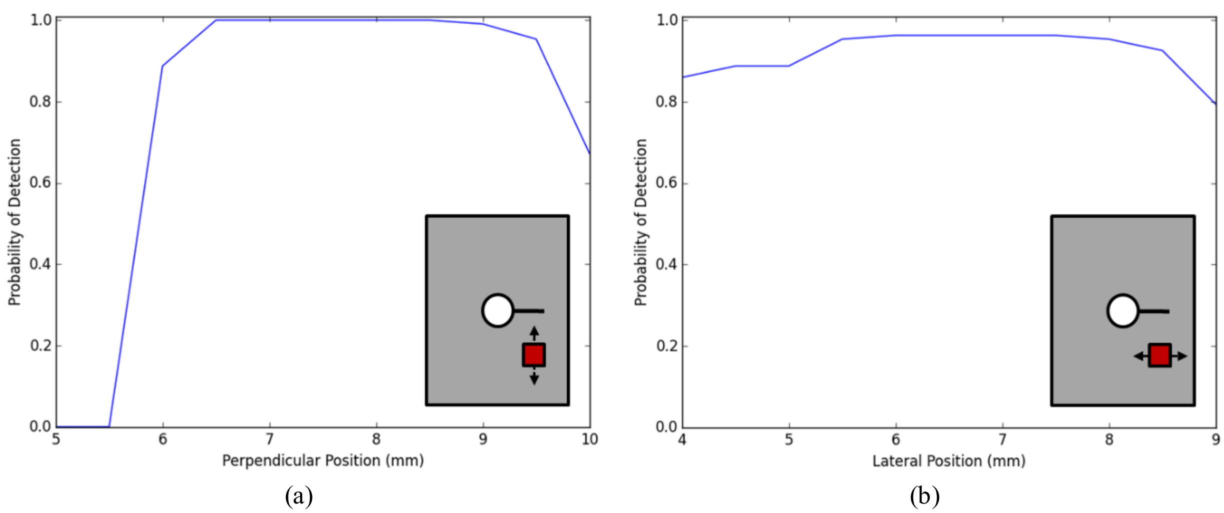


Figure 4. The probability of detection as function of (a) the perpendicular position and (b) the lateral position.

TABLE 1. The ANOVA decomposition of the final fitted model for the two parameter scenario.

Function	Relative Standard Deviation
Perpendicular Position	1
Lateral Position	0.58
Lateral-Perpendicular Interaction	0.74

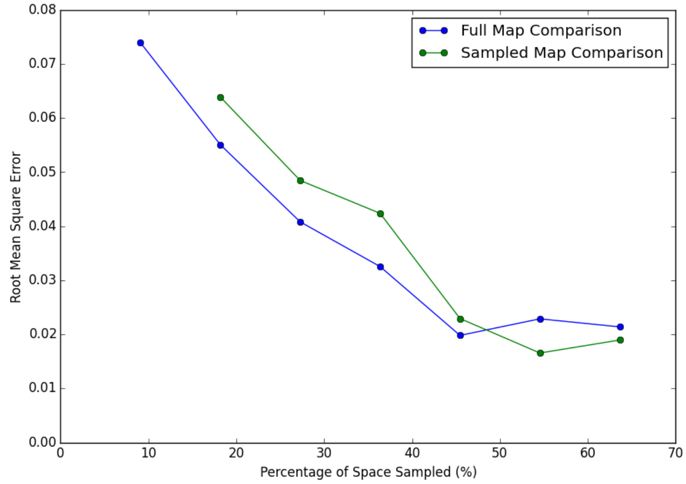


Figure 5. The error of the prediction of the model normalized to the maximum response amplitude, as generated through the mapping algorithm, for the comparison to the sampled response map and the full response map for the two parameter scenario.

Three Parameter Scenario

During inspections performed in-service the operator is likely to rotate the probe on the surface as well as moving it to maximize the response therefore in this scenario the probe is allowed to rotate about the vertical axis as well as moving laterally and perpendicularly. The allowed rotations are in the range [-10°, 10°] and the probability distribution associated with this parameter is a Gaussian distribution with a mean of 0° and a standard deviation of 3°. The same variation ranges and probability distributions were used for the lateral and perpendicular positions as in the two variations scenario. The total parameter space was 1331 points which, given the run time of a single simulation, could not be fully mapped so only the sampling algorithm discussed previously was used.

The ANOVA decomposition of the final fitted model is shown in Table 2. This demonstrates that when the probe is allowed to rotate, the lateral position becomes the most important parameter. It also shows that there is interaction between all of the parameters and that all of the parameters have a significant effect on the response so could not be discounted from the parameter space to reduce the dimensionality. This highlights the importance of considering all variations as discounting some before knowledge of their importance can change the outcome of the qualification result. This information would be very useful for anyone designing an inspection as it highlights what parts of an inspection need greatest attention paid to as well as providing insight into how the inspection may be optimized.

TABLE 2. The ANOVA decomposition of the final fitted model for the three parameter scenario.

Function	Relative Standard Deviation
Perpendicular Position	0.29
Lateral Position	1
Angle	0.5
Lateral-Perpendicular Interaction	0.07
Perpendicular-Angle Interaction	0.51
Lateral-Angle Interaction	0.89
Lateral-Perpendicular-Angle Interaction	0.90

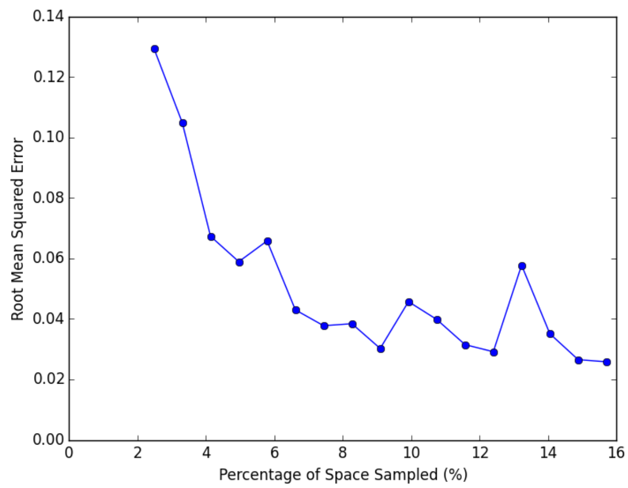


Figure 6. The error of the prediction of the model normalized to the maximum response amplitude, as generated through the mapping algorithm, compared to the sampled response map for the three parameter scenario.

The result of the mapping algorithm is shown in Fig. 6. This shows that a small predictive error is obtained using only a small proportion of the parameter space having been mapped. Compared to the same process applied in the two parameter scenario, the percentage of samples required to obtain a similar predictive error is approximately an order of magnitude smaller. This is indicative of a general trend that these initial results show that as the number of parameters increases, the number of sample points required to accurately map the space increases more slowly than the power law that the number of points in the parameter space follows. Therefore considering more parameters does not necessarily become prohibitive to this process.

CONCLUSIONS

This paper has demonstrated how sampling and interpolation can significantly reduce the number of simulations that have to be performed in order to quantify the reliability of an inspection using models. The combination of Latin Hypercube Designs with Multivariate Adaptive Regression Splines allows the responses of inspections which result in single valued scalar quantities to be mapped for variations present in the inspections using only a small proportion of the parameter space of variations. It is also shown how the output of this process can be used to assess the reliability of an inspection through traditional metrics such as probability of detection. It also yields further information such as relative importance of variations and the interaction between them to be assessed, providing insight into designing and optimizing inspections.

ACKNOWLEDGEMENTS

This work was supported by Dstl, the Royal Commission for the Exhibition of 1851 and the UK Engineering and Physical Sciences Research Council through the UK Research Centre in NDE (RCNDE).

REFERENCES

1. R. A. Smith, "Military Aircraft Structures Airworthiness Advisory Group Paper 122: Development of a protocol for acceptance of new NDT capability in the air domain." Technical Report, Ministry of Defence, 2014.
2. J. C. Helton and F. J. Davis, Reliability Engineering and System Safety **81**, 23-69 (2003).
3. R. Jin, W. Chen and A. Sudjianto, Journal of Statistical Planning and Inference **134 (1)**, 268-287 (2005).
4. F. A. C. Viana, G. Venter and V. Balabanov, International Journal for Numerical Methods in Engineering **82 (2)**, 135-156 (2010).
5. J. H. Friedmann, Annals of Statistics **19 (1)**, 1-141 (1991).
6. P. Huthwaite, Journal of Computational Physics **257**, 687-707 (2014).